Journal of Multivariate Analysis **82**, 65–87 (2002)
doi:10.1006/jmva.2001.2010

Parameter Estimation in Linear Models with Heteroscedastic Variances Subject to Order Restrictions

Carol Hoferkamp

Truman State University, Kirksville, Missouri

and

Shyamal Das Peddada

Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina

Estimation of parameters in linear fixed and mixed effects models, under order restrictions on the error variances, is considered in this article. For simplicity of exposition, we shall assume that the error variances are subject to simple order restriction. Similar methodology can be developed for other forms of order restrictions as well. © 2001 Elsevier Science (USA)

AMS 2000 subject classifications: 62F10; 62F30.

Key words and phrases: fixed effects; heteroscedastic errors; isotonic regression; maximum likelihood estimation; mixed effects; simple order restriction.

1. INTRODUCTION

Suppose data are collected from k independent experimental centers to study the effect of a collection of covariates $\mathbf{X}$ on a response variable Y. For $i = 1, 2, \ldots, k$, let Y_i denote the $n_i \times 1$ response vector observed at the ith center and $\mathbf{X}_i$ denote the corresponding $n_i \times p$ matrix of observed values of p covariates. We shall assume that each $\mathbf{X}_i$ is full rank and denote the total sample size by $N = \sum_{i=1}^{k} n_i$.

Let $Y = [Y_1' : Y_2' : \cdots : Y_k']'$, $\mathbf{X} = [\mathbf{X}_1' : \mathbf{X}_2' : \cdots : \mathbf{X}_k']'$. Assume that Y is a multivariate normal random vector with $E(Y) = \mathbf{X}\beta$ and $\mathrm{Cov}(Y) = \mathbf{\Psi} = \mathbf{\Phi} + \mathbf{\Sigma}$. In the case of fixed effects linear models $\mathbf{\Phi} = 0$, while in the case of mixed effects linear models $\mathbf{\Phi} = \mathbf{U}\mathbf{T}\mathbf{U}'$, where $\mathbf{U}$ is a matrix of known constants and $\mathbf{T}$ is a matrix involving unknown variance components of the random effects. We shall define $\mathbf{U}$ and $\mathbf{T}$ more precisely in

0047-259X/01 $35.00

Section 3. In fixed as well as mixed effects linear models with heteroscedastic errors, $\mathbf{\Sigma} = \mathrm{diag}[\sigma_1^2 \mathbf{I}_{n_1} : \sigma_2^2 \mathbf{I}_{n_2} : \cdots : \sigma_k^2 \mathbf{I}_{n_k}]$, where $\mathrm{diag}[\mathbf{A}_1 : \mathbf{A}_2 : \cdots : \mathbf{A}_k]$ denotes a block-diagonal matrix with diagonal blocks $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$.

An investigator may know *a priori* that some centers may conduct experiments with greater precision than others due to the availability of better technicians and/or better instruments. Thus an experimenter may believe that $\sigma^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)' \in \mathscr{D} \subset \mathscr{R}_+^k$, where $\mathscr{R}_+^k$ is the set of k dimensional positive real numbers. For instance, if the components of σ^2 are subject to *simple order restriction* then

$$\mathscr{D} = \{\sigma^2 \in \mathscr{R}_+^k \mid \sigma_1^2 \leqslant \sigma_2^2 \leqslant \cdots \leqslant \sigma_k^2\}. \tag{1}$$

For simplicity of exposition, in this article we shall consider only the case when σ^2 is subject to simple order restriction. Thus throughout this paper $\mathscr{D}$ is given by (1). In Section 2 we develop methodology for estimating β and $\sigma^2 \in \mathscr{D}$ in fixed effects linear models.

In addition to the estimation of β and $\sigma^2 \in \mathscr{D}$, in mixed effects linear models we need to non-negatively estimate the variance components introduced by the random effects. In Section 3 we develop a new algorithm that uses the EM algorithm together with the methodology introduced in Section 2.

All proofs are provided in the appendix of this article.

2. FIXED EFFECTS MODEL

Shi (1994) and Shi and Jiang (1998) discussed the derivation of restricted maximum likelihood estimators for parameters of k independent normal populations with order restrictions on the means and variances. Although we prove our theorems using some of the techniques developed in these papers, our proofs are complicated by the generality of our model.

The main idea underlying our proposed algorithm, Algorithm 2.1, is to iterate until convergence between the weighted least squares estimator for β and the isotonic regression estimator for σ^2.

ALGORITHM 2.1. Let $\beta^{(m)}$ and $\sigma^{2(m)}$ denote the mth iterate estimates of β and σ^2, respectively. Let $p^{2(m)} = (p_1^{2(m)}, p_2^{2(m)}, \ldots, p_k^{2(m)})'$ and $p_i^{2(m)} = (1/n_i)\|Y_i - \mathbf{X}_i \beta^{(m-1)}\|^2$, where $\|u\|$ indicates the Euclidean norm of vector u defined by $(u'u)^{1/2}$.

Step 0. Compute $\beta^{(0)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$, the ordinary least squares estimator for β. Compute $\sigma_i^{2(0)} = (1/n_i)\,\|Y_i - \mathbf{X}_i \beta^{(0)}\|^2$ for $i = 1, 2, \ldots, k$.

Step 1. Compute the isotonic regression estimator $\sigma^{2(m)}$ by projecting $p^{2(m)}$ onto the cone $\mathscr{D}$ with weights $w=(n_1, n_2, ..., n_k)'$.

Step 2. Compute $\beta^{(m)}=(\mathbf{X}'\mathbf{W}_m\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_m Y$, the weighted least squares estimator of β with weight matrix

$$\mathbf{W}_m=\operatorname{diag}[(1/\sigma_1^{2(m)})\,\mathbf{I}_{n_1} : (1/\sigma_2^{2(m)})\,\mathbf{I}_{n_2} : \cdots : (1/\sigma_k^{2(m)})\,\mathbf{I}_{n_k}].$$

Steps 1 and 2 are iterated until convergence.

We now discuss the convergence of the above algorithm. Let $l(\beta, \sigma^2)$ denote the log-likelihood function under the assumption that the random errors are independent and normally distributed. Suppose $\hat{\beta}_R$ and $\hat{\sigma}_R^2$ denote the maximum likelihood estimators (MLEs) of β and σ^2, respectively, under the restriction $\sigma^2 \in \mathscr{D}$. Then by definition

$$\begin{aligned} l(\hat{\beta}_R, \hat{\sigma}_R^2) &= \sup_{\sigma^2 \in \mathscr{D},\, \beta} l(\beta, \sigma^2) \\ &= \sup_{\sigma^2 \in \mathscr{D}} [L(\sigma^2)], \end{aligned} \tag{2}$$

where $L(\sigma^2)=\sup_\beta l(\beta \mid \sigma^2)$.

Let $\mathbf{W}_\sigma=\operatorname{diag}[(1/\sigma_1^2)\,\mathbf{I}_{n_1} : (1/\sigma_2^2)\,\mathbf{I}_{n_2} : \cdots : (1/\sigma_k^2)\,\mathbf{I}_{n_k}]$. It is well known that the MLE of β, when σ^2 is known, is given by $(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y$. Therefore

$$l(\beta^{(m-1)}, \sigma^{2(m)}) \leqslant l(\beta^{(m)}, \sigma^{2(m)}). \tag{3}$$

Additionally, under the order restriction given by $\mathscr{D}$, the isotonic regression of $\sigma^{2(m)}$ is the MLE of σ^2 when $\beta^{(m)}$ is known (Robertson *et al.*, 1988). Hence

$$l(\beta^{(m)}, \sigma^{2(m)}) \leqslant l(\beta^{(m)}, \sigma^{2(m+1)}). \tag{4}$$

Thus at each step of the algorithm the likelihood is increased.

DEFINITION 2.1. Define a favorable point $\sigma^2 \in \mathscr{D}$ as one for which there exists a subscript set $\{i_1, i_2, ..., i_t\}$ with $1 \leqslant i_1 < i_2 < \cdots < i_t < k$ such that

$$\sigma_1^2 = \cdots = \sigma_{i_1}^2 < \sigma_{i_1+1}^2 = \cdots = \sigma_{i_2}^2 < \cdots < \sigma_{i_t+1}^2 = \cdots = \sigma_k^2 \tag{5}$$

and

$$\sum_{i=i_s+1}^{i_{(s+1)}} \{n_i\sigma_i^2 - \|Y_i - \mathbf{X}_i(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y\|^2\}=0 \tag{6}$$

for $s=0, 1, ..., t$ where $i_0=0$ and $i_{(t+1)}=k$.

THEOREM 2.1. *The MLE $\hat{\sigma}^2_R$ is a favorable point.*

We now have the following theorem regarding Algorithm 2.1.

THEOREM 2.2. *Let $\{\sigma^{2(m)}\}$ be the sequence of estimated variances from Algorithm 2.1. If there are finitely many favorable points, then $\{\sigma^{2(m)}\}$ converges to a favorable point as $m \to \infty$.*

Verifying that a particular model will have only finitely many favorable points should not be ignored. In the following example, we demonstrate that this condition will be satisfied for a replicated model—where the linear model is replicated k times so that $\mathbf{X}_i \equiv \mathbf{X}_0$, and $n_i \equiv n_0$, for $i = 1, 2, \ldots, k$. Replicated models have been well studied in the literature by many researchers. These models arise very naturally in fertilizer trials where the agronomist may want to study the repeatability of dose responses from year to year. Some useful references in this context are Khosla *et al.* (1979), Rao *et al.* (1987), Rao *et al.* (1998), and Srivastava and Toutenburg (1994).

EXAMPLE 2.1. In the case of the replicated model, we observe that (5) and (6) imply that, for $s = 0, 1, \ldots, t$

$$0 = \theta_s \left[\sum_{u=0}^{t} (a_s - 2b_{su} + a_u)\, \theta_u \right] - \theta_s n(t+1) - n \left(\sum_{u=0}^{t} \theta_u \right), \tag{7}$$

where $\theta_s = 1/\sigma^2_{i_{(s+1)}}$, $a_s = Y_s' Y_s$, $b_{uv} = Y_u' \mathbf{X}_0 (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' Y_v$, and $n = \sum_{i=i_s+1}^{i_{(s+1)}} n_i$. The derivation of (7) can be found in the Appendix.

Observe that (7) defines a system of $t+1$ equations which are quadratic in $\theta_0, \ldots, \theta_t$. From Huber and Sturmfels (1995), a system of k polynomial equations of degree m in k variables has finitely many solutions if all k *facial resultants* are non-zero. Here facial resultants are determinants involving the coefficients of various terms of the polynomial equations (cf. Huber and Sturmfels, 1995; Pedersen and Sturmfels, 1993; Sturmfels, 1994). In (7) the coefficients are functions of the normal random vector Y, which is a continuous random variable. Hence the determinants are non-zero with probability 1. Thus the number of solutions is finite with probability 1.

We thank Professor Sturmfels, Department of Mathematics, University of California at Berkeley for the above justification using his recent results on algebraic geometry.

Thus we have the following theorem.

THEOREM 2.3. *The number of favorable points for a replicated linear model is finite with probability* 1.

Clearly if there is only one favorable point then by virtue of Theorem 1.2 and Theorem 2.1, $\{\sigma^{2(m)}\}$ converges to the MLE as $m \to \infty$.

3. MIXED EFFECTS MODELS

In this section we shall consider the following mixed effects linear model with heteroscedastic errors,

$$Y = \mathbf{X}\beta + \sum_{i=1}^{q} \mathbf{U}_i \xi_i + \varepsilon, \tag{8}$$

where Y is a $N \times 1$ response vector, $\mathbf{X}, \mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_q$ are known design matrices of orders $N \times p$, $N \times c_1$, ..., $N \times c_q$, respectively. Further, ξ_i are independent random effects with mean 0 and covariance matrix $\tau_i^2 \mathbf{I}_{c_i}$, and ε is a $N \times 1$ random error vector with mean 0 and covariance matrix $\mathbf{\Sigma} = \text{diag}[\sigma_1^2 \mathbf{I}_{n_1} : \cdots : \sigma_k^2 \mathbf{I}_{n_k}]$. We shall denote $\tau^2 = (\tau_1^2, \tau_2^2, \ldots, \tau_q^2)'$, $\mathbf{T} = \text{Cov}(\xi_1' : \xi_2' : \ldots : \xi_q')$, and $\mathbf{U} = [\mathbf{U}_1 : \mathbf{U}_2 : \ldots : \mathbf{U}_q]$. Let $\text{tr}(\mathbf{A})$ denote the trace of matrix $\mathbf{A}$.

We estimate β, $\tau^2 \in \mathscr{R}_+^q$, and $\sigma^2 \in \mathscr{D}$ using the following algorithm. The algorithm iterates, until convergence, between estimation of β and τ^2 based on the previous estimate of σ^2 and the estimation of σ^2 based on the previous estimates of β and τ^2. At each step the estimation procedure uses the EM algorithm of Dempster *et al.* (1977) to estimate all the parameters of the model with no restrictions on σ^2 and projects the estimates of σ^2 onto $\mathscr{D}$ using isotonic regression.

ALGORITHM 3.1. Let $\beta^{(m)}$, $\sigma^{2(m)}$, and $\tau^{2(m)}$ denote the mth iterate estimates of β, σ^2, and τ^2 respectively.

Step 0. Set $m = 0$. Compute $\beta^{(0)} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'Y$, the ordinary least squares estimator for β. Compute $\sigma_i^{2(0)} = (1/n_i) \|Y_i - \mathbf{X}_i \beta^{(0)}\|^2$ for $i = 1, 2, \ldots, k$. For $\tau_i^{2(0)}$ we shall use some non-negative, initial estimator such as the MINQE (Rao and Chaubey, 1978).

Step 1. Set $m = m + 1$. In this step we fix β and τ^2 at $\beta^{(m-1)}$ and $\tau^{2(m-1)}$, respectively, and iteratively estimate σ^2 using the following iteration equation derived from the EM algorithm (Theorem 3.1). For $i = 1, 2, \ldots, k$, and $r = 1, \ldots,$

$$\begin{aligned}\hat{\sigma}_i^{2(r)} = \hat{\sigma}_i^{2(r-1)} + (\hat{\sigma}_i^{4(r-1)}/n_i)\, \text{tr}[(\hat{\mathbf{\Psi}}^{(r-1)})^{-1} (Y - \mathbf{X}\beta)(Y - \mathbf{X}\beta)' (\hat{\mathbf{\Psi}}^{(r-1)})^{-1} \\ - (\hat{\mathbf{\Psi}}^{(r-1)})^{-1}]_{ii},\end{aligned} \tag{9}$$

where $(\mathbf{A})_{ii}$ indicates the (i, i) block of matrix $\mathbf{A}$, $\hat{\sigma}^{2(0)} = \sigma^{2(m-1)}$, and $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{UTU}' + \hat{\boldsymbol{\Sigma}}^{(r-1)}$. Let $\hat{\sigma}^{2(r^*)}$ be the solution obtained at the end of the iteration process described in (9). Compute the isotonic regression estimate $\sigma^{2(m)}$ by projecting $\hat{\sigma}^{2(r^*)}$ onto the cone $\mathscr{D}$ with weight vector $w = (n_1, n_2, \ldots, n_k)'$.

Step 2. Fixing σ^2 at $\sigma^{2(m)}$, we iteratively estimate β and τ^2 using the following simultaneous iteration equations derived from the EM algorithm (Theorem 3.2). For $i = 1, 2, \ldots, q$, and $r = 1, \ldots,$

$$\hat{\beta}^{(r)} = \hat{\beta}^{(r-1)} + (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\, \mathbf{X}'(\hat{\boldsymbol{\Psi}}^{(r-1)})^{-1}\, (Y - \mathbf{X}\hat{\beta}^{(r-1)}), \tag{10}$$

and

$$\begin{aligned} \hat{\tau}_i^{2(r)} &= \hat{\tau}_i^{2(r-1)} + (\hat{\tau}_i^{4(r-1)}/c_i) \\ &\quad \cdot \operatorname{tr} \mathbf{U}_i'[(\hat{\boldsymbol{\Psi}}^{(r-1)})^{-1}\, (Y - \mathbf{X}\hat{\beta}^{(r-1)})(Y - \mathbf{X}\hat{\beta}^{(r-1)})'\, (\hat{\boldsymbol{\Psi}}^{(r-1)})^{-1} \\ &\quad - (\hat{\boldsymbol{\Psi}}^{(r-1)})^{-1}]\, \mathbf{U}_i, \end{aligned} \tag{11}$$

where $\hat{\beta}^{(0)} = \beta^{(m-1)}$, $\hat{\tau}^{2(0)} = \tau^{2(m-1)}$, and $\hat{\boldsymbol{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \boldsymbol{\Sigma}$. Let $\beta^{(m)}$ and $\tau^{2(m)}$ be the solutions obtained at the end of the iteration process described by (10) and (11), respectively.

Steps 1 and 2 are iterated until convergence.

As in Section 2 it is easy to verify that the likelihood increases at each step of the algorithm. Due to the convergence of the EM algorithm we conjecture that Algorithm 3.1 converges to a favorable point. The proof of this statement will require results analogous to those derived in Section 2. As in the case of the fixed effects models, such results are likely to be very hard to prove and they will depend upon the covariance structure of the mixed effects model.

The following theorems derive the iterative equations for the EM algorithm.

THEOREM 3.1. *The EM estimates, at the r th iteration, for σ^2 when β and τ^2 are known are given by* (9).

THEOREM 3.2. *The EM estimates, at the r th iteration, for β and τ^2 when σ^2 is known are given by* (10) *and* (11) *respectively.*

4. CONCLUSIONS AND OPEN PROBLEMS

In this article we consider the estimation of parameters in linear models, under order restrictions on the error variances. For the fixed effects model, we propose an algorithm which iterates between the weighted least squares estimator for the regression parameter and the isotonic regression estimator for the variances. For the mixed effects model, the proposed algorithm uses the EM algorithm to estimate all the parameters of the model with no restrictions on the variances and projects the estimates of σ^2 onto $\mathscr{D}$ using isotonic regression. While we assumed in the exposition that the error variances are subject to simple order restriction, similar methodology can be developed for other order restriction patterns as well.

In the case of fixed effects models, we have established that the proposed algorithm converges to a favorable point, assuming that the number of favorable points is finite. We have shown that this condition holds in the case of replicated models, but the issue of finiteness for a general linear model remains unsettled. In the case of mixed effects models, we conjecture that the algorithm may also converge to a favorable point. However, the proof of such a result is nontrivial and will depend on the covariance structure of the variance component model and the convergence of the EM algorithm. Finally, it remains to study the statistical properties of the estimators (e.g. domination in terms of MSE) that result from these algorithms. This is an important issue that needs to be addressed in the literature. Surprisingly, to the best of our knowledge, such domination results do not exist even for the simple case of estimating μ_i and σ_i^2 in the model $Y_i \sim^{independent} N(\mu_i, \sigma_i^2)$, with order restrictions on μ_i's and σ_i^2's. Results in Hwang and Peddada (1994) are obtained for the case when the variances are known but means are unknown with order restrictions on them. We believe this a very rich area for future research.

APPENDIX

Proof of Theorem 2.1. We will now show that $\hat{\sigma}_R^2$ is a favorable point. Observe from (2) that $l(\hat{\beta}_R, \hat{\sigma}_R^2) = \sup_{\sigma^2 \in \mathscr{D}} [L(\sigma^2)]$. Thus $\hat{\sigma}_R^2$ maximizes the Lagrangian function

$$\mathbf{\Phi}(\sigma^2, \lambda) = L(\sigma^2) + \sum_{i=1}^{k-1} \lambda_i(\sigma_{i+1}^2 - \sigma_i^2) \qquad \text{subject to} \quad \sigma^2 \in \mathscr{D},$$

where λ_i are the Lagrangian multipliers. According to the Kuhn–Tucker conditions, $\hat{\sigma}_R^2$ satisfies

(i) $\hat{\sigma}_R^2 \in \mathscr{D}$

(ii) $\left.\dfrac{\partial \mathbf{\Phi}}{\partial \sigma_i^2}\right|_{\hat{\sigma}_{R(i)}^2} = 0 \qquad i = 1, \ldots, k$

(iii) $\lambda_i \geqslant 0 \qquad i = 1, \ldots, k-1$

(iv) $\lambda_i(\hat{\sigma}_{R(i+1)}^2 - \hat{\sigma}_{R(i)}^2) = 0 \qquad i = 1, \ldots, k-1.$

Let $\{i_1, i_2, \ldots, i_t\}$ be the subscript set such that $\hat{\sigma}^2_{R(i+1)} > \hat{\sigma}^2_{R(i)}$ for $i \in \{i_1, i_2, \ldots, i_t\}$ and $\hat{\sigma}^2_{R(i+1)} = \hat{\sigma}^2_{R(i)}$ otherwise. From (i) we note that there exists an index set such that (5) holds.

For simplicity of notation, denote $\beta_\sigma = (\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y$. Observe that $L(\sigma^2) = \sup_\beta l(\beta \mid \sigma^2) = l(\beta_\sigma \mid \sigma^2)$. Based on routine calculations we have

$$
\begin{aligned}
\frac{\partial L(\sigma^2)}{\partial \sigma_i^2} &= \frac{\partial}{\partial \sigma_i^2}\Bigg\{\text{const} + \sum_{i=1}^{k} [-(n_i/2)\ln \sigma_i^2 \\
&\quad - (1/2)(1/\sigma_i^2)(Y_i - \mathbf{X}_i\beta_\sigma)'\,(Y_i - \mathbf{X}_i\beta_\sigma)]\Bigg\} \\
&= -(n_i/2)(1/\sigma_i^2) - (1/2)\frac{\partial}{\partial \sigma_i^2}\{(Y - \mathbf{X}(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y)'\,\mathbf{W}_\sigma \\
&\quad \cdot (Y - \mathbf{X}(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y)\} \\
&= -(n_i/2)(1/\sigma_i^2) - (1/2)\frac{\partial}{\partial \sigma_i^2}(Y'\mathbf{W}_\sigma Y - Y'\mathbf{W}_\sigma\mathbf{X}(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y).
\end{aligned}
$$

Recall that

$$
\frac{\partial \mathbf{B}^{-1}}{\partial x} = -\mathbf{B}^{-1}\left(\frac{\partial \mathbf{B}}{\partial x}\right)\mathbf{B}^{-1},
$$

therefore

$$
\frac{\partial (\mathbf{A}'\mathbf{B}^{-1}\mathbf{A})^{-1}}{\partial x} = (\mathbf{A}'\mathbf{B}^{-1}\mathbf{A})^{-1}\,\mathbf{A}'\mathbf{B}^{-1}\left(\frac{\partial \mathbf{B}}{\partial x}\right)\mathbf{B}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{B}^{-1}\mathbf{A})^{-1},
$$

where $\mathbf{A}$ does not depend on x. Since $Y'\mathbf{W}_\sigma Y = \sum (1/\sigma_i^2)\,Y_i'Y_i$, it follows that

$$
\begin{aligned}
\frac{\partial L(\sigma^2)}{\partial \sigma_i^2} &= (1/2)(1/\sigma_i^4)[-n_i\sigma_i^2 + Y_i'Y_i - Y_i'\mathbf{X}_i(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y \\
&\quad + Y'\mathbf{W}_\sigma\mathbf{X}(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_\sigma Y \\
&\quad - Y'\mathbf{W}_\sigma\mathbf{X}(\mathbf{X}'\mathbf{W}_\sigma\mathbf{X})^{-1}\,\mathbf{X}_i'Y_i] \\
&= -(1/2)(1/\sigma_i^4)[n_i\sigma_i^2 - \|Y_i - \mathbf{X}_i\beta_\sigma\|^2].
\end{aligned}
$$

Notice that $\partial\mathbf{\Phi}/\partial\sigma_i^2 = (\partial L(\sigma^2)/\partial\sigma_i^2) + \lambda_{i-1} - \lambda_i$. Therefore, (ii) is equivalent to

$$
-(1/2)(1/\hat{\sigma}^4_{R(i)})[n_i\hat{\sigma}^2_{R(i)} - \|Y_i - \mathbf{X}_i\beta_{\hat{\sigma}_R}\|^2] + \lambda_{i-1} - \lambda_i = 0,
$$

or equivalently,

$$n_i \hat{\sigma}^2_{R(i)} - \|Y_i - \mathbf{X}_i \beta_{\hat{\sigma}_R}\|^2 = 2(\lambda_{i-1} - \lambda_i)\, \hat{\sigma}^4_{R(i)}.$$

Taking the sum of each side above from $i_s + 1$ to $i_{(s+1)}$, then (ii) implies that

$$\begin{aligned}\sum_{i=i_s+1}^{i_{(s+1)}} (n_i \hat{\sigma}^2_{R(i)} - \|Y_i - \mathbf{X}_i \beta_{\hat{\sigma}_R}\|^2) &= 2 \sum_{i=i_s+1}^{i_{(s+1)}} (\lambda_{i-1} - \lambda_i)\, \hat{\sigma}^4_{R(i)} \\ &= 2\hat{\sigma}^4_{R(i_{(s+1)})} \sum_{i=i_s+1}^{i_{(s+1)}} (\lambda_{i-1} - \lambda_i) \qquad \text{by (5)} \\ &= 2\hat{\sigma}^4_{R(i_{(s+1)})} (\lambda_{i_s} - \lambda_{i_{(s+1)}}).\end{aligned}$$

By the choice of the index set, $\hat{\sigma}^2_{R(i_s+1)} > \hat{\sigma}^2_{R(i_s)}$ and $\hat{\sigma}^2_{R(i_{(s+1)}+1)} > \hat{\sigma}^2_{R(i_{(s+1)})}$. Hence by virtue of (iv), $\lambda_{i_s} = 0$ and $\lambda_{i_{(s+1)}} = 0$. Therefore the left hand side above is 0. Thus (6) is satisfied for $s = 0, 1, \ldots, t$, and the theorem is proved.

Derivation of (7) *for Example* 2.1. We begin with a general observation from the definition of a favorable point and then apply this observation to the case of the replicated model.

Consider a favorable point σ^2 with a fixed subscript set according to (5). For $i \in \{i_s + 1, i_s + 2, \ldots, i_{(s+1)}\}$, $s = 0, 1, \ldots, t$ where $i_0 = 0$ and $i_{(t+1)} = k$, then σ^2_i is a constant. From (6), then

$$\begin{aligned}0 &= \sum_{i=i_s+1}^{i_{(s+1)}} [n_i \sigma^2_{i_{(s+1)}} - \|Y_i - \mathbf{X}_i (\mathbf{X}'\mathbf{W}_\sigma \mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_\sigma Y\|^2] \\ &= \left(\sum_{i=i_s+1}^{i_{(s+1)}} n_i\right) \sigma^2_{i_{(s+1)}} - \sum_{i=i_s+1}^{i_{(s+1)}} \|Y_i - \mathbf{X}_i (\mathbf{X}'\mathbf{W}_\sigma \mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_\sigma Y\|^2 \\ &= \left(\sum_{i=i_s+1}^{i_{(s+1)}} n_i\right) \sigma^2_{i_{(s+1)}} - \|Y_s - \mathbf{X}_s (\mathbf{X}'\mathbf{W}_\sigma \mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_\sigma Y\|^2, \end{aligned} \tag{12}$$

where $Y_s = [Y'_{i_s+1} : \cdots : Y'_{i_{(s+1)}}]'$ and $\mathbf{X}_s = [\mathbf{X}'_{i_s+1} : \cdots : \mathbf{X}'_{i_{(s+1)}}]'$ for $s = 0, 1, \ldots, t$. Observe that there are finitely many subscripts sets, as $k < \infty$. Thus to show that there are finitely many favorable points, it is sufficient to show that the system of equations defined by (12) has a finite number of solutions.

We now consider the simplification of (12) in the case of the replicated model. Define $\theta_s = 1/\sigma^2_{i_{(s+1)}}$ for $s = 0, 1, \ldots, t$. Notice that for the replicated model, $\mathbf{X}'\mathbf{W}_\sigma \mathbf{X} = \sum_{u=0}^{t} \theta_u \mathbf{X}'_0 \mathbf{X}_0$ and so $(\mathbf{X}'\mathbf{W}_\sigma \mathbf{X})^{-1} = (\frac{1}{\sum_{u=0}^{t} \theta_u})\, (\mathbf{X}'_0 \mathbf{X}_0)^{-1}$.

Note that

$$\begin{aligned}
&\|Y_s - \mathbf{X}_s(\mathbf{X}'\mathbf{W}_\sigma \mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_\sigma Y\|^2\\
&\quad = \left\| Y_s - \mathbf{X}_0\left(\frac{1}{\sum_{u=0}^t \theta_u}\right)(\mathbf{X}_0'\mathbf{X}_0)^{-1}\left(\sum_{u=0}^t \theta_u \mathbf{X}_0' Y_u\right)\right\|^2\\
&\quad = \left\| Y_s - \left(\frac{\sum_{u=0}^t \theta_u \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0' Y_u}{\sum_{u=0}^t \theta_u}\right)\right\|^2\\
&\quad = Y_s'Y_s - 2\left(\frac{\sum_{u=0}^t \theta_u Y_s'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0' Y_u}{\sum_{u=0}^t \theta_u}\right)\\
&\qquad + \left(\frac{\sum_{u=0}^t \sum_{v=0}^t \theta_u\theta_v Y_u'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0' Y_v}{(\sum_{u=0}^t \theta_u)^2}\right).
\end{aligned}$$

To simplify notation, let $a_s = Y_s'Y_s$, $b_{uv} = Y_u'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'Y_v$, and $n = \sum_{i=i_s+1}^{i_{(s+1)}} n_i$. Then (12) becomes

$$0 = n\theta_s^{-1} - \left[a_s - 2\left(\frac{\sum_{u=0}^t \theta_u b_{su}}{\sum_{u=0}^t \theta_u}\right) + \frac{\sum_{u=0}^t \sum_{v=0}^t \theta_u\theta_v b_{uv}}{(\sum_{u=0}^t \theta_u)^2}\right]$$

or equivalently

$$\begin{aligned}
0 &= a_s\theta_s\left(\sum_{u=0}^t \theta_u\right)^2 - 2\theta_s\left(\sum_{u=0}^t \theta_u\right)\sum_{u=0}^t \theta_u b_{su}\\
&\quad + \theta_s \sum_{u=0}^t \sum_{v=0}^t \theta_u\theta_v b_{uv} - n\left(\sum_{u=0}^t \theta_u\right)^2,
\end{aligned} \tag{13}$$

for $s = 0, 1, \ldots, t$. Taking the sum of the $t+1$ equations defined by (13) we have

$$\begin{aligned}
0 &= \left(\sum_{s=0}^t a_s\theta_s\right)\left(\sum_{u=0}^t \theta_u\right)^2 - 2\left(\sum_{u=0}^t \theta_u\right)\sum_{s=0}^t \theta_s \sum_{u=0}^t \theta_u b_{su}\\
&\quad + \left(\sum_{s=0}^t \theta_s\right)\sum_{u=0}^t \sum_{v=0}^t \theta_u\theta_v b_{uv} - n(t+1)\left(\sum_{u=0}^t \theta_u\right)^2\\
&= \left(\sum_{u=0}^t a_u\theta_u\right)\left(\sum_{u=0}^t \theta_u\right)^2 - \left(\sum_{u=0}^t \theta_u\right)\sum_{u=0}^t \sum_{v=0}^t \theta_u\theta_v b_{uv} - n(t+1)\left(\sum_{u=0}^t \theta_u\right)^2.
\end{aligned}$$

Dividing through by $\sum_{u=0}^t \theta_u$, as $\sum_{u=0}^t \theta_u \neq 0$, the above expression simplifies as

$$\sum_{u=0}^t \sum_{v=0}^t \theta_u\theta_v b_{uv} = \left(\sum_{u=0}^t a_u\theta_u\right)\left(\sum_{u=0}^t \theta_u\right) - n(t+1)\left(\sum_{u=0}^t \theta_u\right). \tag{14}$$

Substituting (14) into each of the equations from (13) yields for $s = 0, 1, \ldots, t$

$$0 = a_s\theta_s\left(\sum_{u=0}^{t}\theta_u\right)^2 - 2\theta_s\left(\sum_{u=0}^{t}\theta_u\right)\sum_{u=0}^{t}\theta_u b_{su} + \theta_s\left(\sum_{u=0}^{t}a_u\theta_u\right)\left(\sum_{u=0}^{t}\theta_u\right)$$
$$-\theta_s n(t+1)\left(\sum_{u=0}^{t}\theta_u\right) - n\left(\sum_{u=0}^{t}\theta_u\right)^2.$$

Again we observe that $\sum_{u=0}^{t}\theta_u \neq 0$; thus the above expression reduces to

$$0 = a_s\theta_s\left(\sum_{u=0}^{t}\theta_u\right) - 2\theta_s\sum_{u=0}^{t}\theta_u b_{su} + \theta_s\left(\sum_{u=0}^{t}a_u\theta_u\right) - \theta_s n(t+1) - n\left(\sum_{u=0}^{t}\theta_u\right),$$

for $s = 0, 1, \ldots, t$. Minor simplification of the above equation gives (7).

The following lemmas will be used in proving Theorem 2.2.

LEMMA A.1. *Let $\{\sigma^{2(m)}\}$ be the sequence of estimated variances from Algorithm 2.1. If $0 < (1/n_i)\,\|Y_i - \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\,\mathbf{X}_i'Y_i\|^2$ for all $i = 1, 2, \ldots, k$ and if $Y'Y < \infty$, then $0 < \sigma_i^{2(m)} \leqslant B$ where B is some finite constant that does not depend on i nor m. In other words, $\{\sigma^{2(m)}\}$ is a uniformly bounded sequence.*

Proof. Since $\sigma^{2(m)}$ is the isotonic regression of $p^{2(m)}$ with weights w, therefore for all $i = 1, 2, \ldots, k$,

$$\min_{i=1,2,\ldots,k}\{p_i^{2(m)}\} \leqslant \sigma_i^{2(m)} \leqslant \max_{i=1,2,\ldots,k}\{p_i^{2(m)}\}. \tag{15}$$

Note that

$$p_i^{2(m)} = (1/n_i)\,\|Y_i - \mathbf{X}_i(\mathbf{X}'\mathbf{W}_{(m-1)}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_{(m-1)}Y\|^2$$
$$\geqslant (1/n_i)\,\|Y_i - \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\,\mathbf{X}_i'Y_i\|^2 > 0.$$

Hence $0 < \sigma_i^{2(m)}$ for all $i = 1, 2, \ldots, k$ and $m = 1, 2, \ldots$.

Using mathematical induction, we first demonstrate that the $p_i^{2(m)}$ are uniformly upper bounded by $Y'Y$ for all $i = 1, 2, \ldots, k$ and $m = 1, 2, \ldots$. The required result can then be proved by appealing to (15).

Observe that since $n_i \geqslant 1$, then

$$\begin{aligned} p_i^{2(m)} &\leqslant n_i(1/n_i)\,\|Y_i - \mathbf{X}_i\beta^{(m-1)}\|^2 \\ &= \|Y_i - \mathbf{X}_i\beta^{(m-1)}\|^2 \\ &\leqslant \sum_{i=1}^{k}\|Y_i - \mathbf{X}_i\beta^{(m-1)}\|^2 \\ &= \|Y - \mathbf{X}\beta^{(m-1)}\|^2. \end{aligned} \tag{16}$$

In particular, $p_i^{2(1)} \leqslant \|Y - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y\|^2$. Since $\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$ minimizes $\|Y - \mathbf{X}\beta\|^2$, then $p_i^{2(1)} \leqslant \|Y - \mathbf{X}0\|^2 = Y'Y$. Thus $p_i^{2(1)} \leqslant Y'Y$.

Now suppose that $0 < p_i^{2(m-1)} \leqslant Y'Y$ for $i = 1, 2, ..., k$. From (15) then $0 < \sigma_i^{2(m-1)} < \infty$. Hence $\mathbf{W}_{(m-1)}$ is a positive definite matrix. To simplify notation, let $\mathbf{W} = \mathbf{W}_{(m-1)}$. Let $\mathbf{U} = \mathbf{W}^{(1/2)}\mathbf{X}$ and $\mathbf{X} = \mathbf{W}^{-(1/2)}\mathbf{U}$. Then

$$\begin{aligned} \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{W}_{(m-1)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_{(m-1)} &= \mathbf{I}_N - \mathbf{W}^{-(1/2)}\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{W}^{(1/2)} \\ &= \mathbf{W}^{-(1/2)}[\mathbf{I}_N - \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}']\,\mathbf{W}^{(1/2)}. \end{aligned}$$

Let $\mathbf{Q} = \mathbf{I}_N - \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'$ and $\lambda_{\max}(\mathbf{A})$ be the largest eigenvalue of $\mathbf{A}$. Consider that

$$\begin{aligned} \lambda_{\max}(\mathbf{W}^{(1/2)}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}^{(1/2)}) &= \lambda_{\max}(\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}) \\ &\leqslant \lambda_{\max}(\mathbf{Q})\,\lambda_{\max}(\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}) \\ &= \lambda_{\max}(\mathbf{Q})\,\lambda_{\max}(\mathbf{Q}\mathbf{W}\mathbf{W}^{-1}) \\ &= \lambda_{\max}(\mathbf{Q})\,\lambda_{\max}(\mathbf{Q}) = 1, \end{aligned}$$

as $\mathbf{Q}$ is a projection matrix and has eigenvalues of only 1 and 0. Let $\lambda_{\min}(\mathbf{A})$ be the smallest eigenvalue of $\mathbf{A}$. Since

$$\begin{aligned} \lambda_{\min}(\mathbf{I}_N - \mathbf{W}^{(1/2)}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}^{(1/2)}) &= 1 - \lambda_{\max}(\mathbf{W}^{(1/2)}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}^{(1/2)}) \\ &\geqslant 1 - 1 = 0, \end{aligned}$$

it follows that $\mathbf{I}_N - \mathbf{W}^{(1/2)}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}^{(1/2)}$ is a non-negative definite matrix. Thus

$$Y'[\mathbf{I}_N - \mathbf{W}^{(1/2)}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}^{(1/2)}]\,Y \geqslant 0.$$

Note that

$$\begin{aligned} Y'[\mathbf{I}_N - \mathbf{W}^{(1/2)}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}^{(1/2)}]\,Y &= Y'Y - Y'\mathbf{W}^{(1/2)}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}\mathbf{W}^{(1/2)}Y \\ &= Y'Y - \|Y - \mathbf{X}\beta^{(m-1)}\|^2. \end{aligned}$$

In other words, $Y'Y \geqslant \|Y - \mathbf{X}\beta^{(m-1)}\|^2$. Therefore, from (16), $p_i^{2(m)} \leqslant Y'Y < \infty$ for all $i = 1, 2, ..., k$ and $m = 1, 2, ...$ Therefore the lemma is proved and $B = Y'Y$.

LEMMA A.2. *Let* $\{\sigma^{2(m)}\}$ *be the sequence of estimated variances from Algorithm* 2.1. *Let* $\{\sigma^{2(m_j)}\}$ *be a subsequence of* $\{\sigma^{2(m)}\}$. *If* $\{\sigma^{2(m_j)}\}$ *is convergent, then* $\sigma_i^{2(m_j)} - \sigma_i^{2(m_j-1)} \to 0$ *for* $i = 1, 2, ..., k$ *as* $m_j \to \infty$.

Proof. Let $\{\beta^{(m_j)}\}$ and $\{l(\beta^{(m_j)}, \sigma^{2(m_j)})\}$ be the corresponding subsequences of regression coefficients and log-likelihoods, respectively, from the algorithm. Since $\{\sigma^{2(m_j)}\}$ converges, and $\beta^{(m_j)}$ is a continuous function of $\sigma^{2(m_j)}$, then $\{\beta^{(m_j)}\}$ is also convergent. Similarly, $\{l(\beta^{(m_j)}, \sigma^{2(m_j)})\}$ is convergent. Hence $l(\beta^{(m_{(j-1)})}, \sigma^{2(m_{(j-1)})}) - l(\beta^{(m_j)}, \sigma^{2(m_j)}) \to 0$ as $m_j \to \infty$. Therefore from (3) and (4) we have

$$l(\beta^{(m_{(j-1)})}, \sigma^{2(m_{(j-1)})}) - l(\beta^{(m_j)}, \sigma^{2(m_j)}) \leqslant l(\beta^{(m_{(j-1)})}, \sigma^{2(m_{(j-1)})}) - l(\beta^{(m_j)}, \sigma^{2(m_j-1)})$$
$$\leqslant l(\beta^{(m_j-1)}, \sigma^{2(m_j-1)}) - l(\beta^{(m_j)}, \sigma^{2(m_j-1)})$$

since $m_{(j-1)} \leqslant m_j - 1$. Since the left hand side converges to 0 and the right hand side is non-positive by (3), then $l(\beta^{(m_j-1)}, \sigma^{2(m_j-1)}) - l(\beta^{(m_j)}, \sigma^{2(m_j-1)})$ must also converge to 0 as $m_j \to \infty$.

Note that

$$\begin{aligned}
&l(\beta^{(m_j-1)}, \sigma^{2(m_j-1)}) - l(\beta^{(m_j)}, \sigma^{2(m_j-1)}) \\
&\quad = \text{const} + \sum_{i=1}^{k} \{-(n_i/2) \ln \sigma_i^{2(m_j-1)} \\
&\qquad - (1/2)(1/\sigma_i^{2(m_j-1)})(Y_i - \mathbf{X}_i\beta^{(m_j-1)})'\,(Y_i - \mathbf{X}_i\beta^{(m_j-1)})\} \\
&\qquad - \Bigg[\text{const} + \sum_{i=1}^{k} \{-(n_i/2) \ln \sigma_i^{2(m_j-1)} \\
&\qquad - (1/2)(1/\sigma_i^{2(m_j-1)})(Y_i - \mathbf{X}_i\beta^{(m_j)})'\,(Y_i - \mathbf{X}_i\beta^{(m_j)})\}\Bigg] \\
&\quad = (1/2) \sum_{i=1}^{k} \{(1/\sigma_i^{2(m_j-1)})(Y_i - \mathbf{X}_i\beta^{(m_j)})'\,(Y_i - \mathbf{X}_i\beta^{(m_j)})\} \\
&\qquad - (1/2) \sum_{i=1}^{k} \{(1/\sigma_i^{2(m_j-1)})(Y_i - \mathbf{X}_i\beta^{(m_j-1)})' \cdot (Y_i - \mathbf{X}_i\beta^{(m_j-1)})\} \\
&\quad = (1/2)\{\|Y - \mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}} - \|Y - \mathbf{X}\beta^{(m_j-1)}\|^2_{\mathbf{W}_{(m_j-1)}}\},
\end{aligned}$$

where $\|u\|_A^2 = u'Au$. Since the left hand side converges to zero, then

$$\|Y - \mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}} - \|Y - \mathbf{X}\beta^{(m_j-1)}\|^2_{\mathbf{W}_{(m_j-1)}} \to 0.$$

Using the fact that $\beta^{(m_j-1)} = (\mathbf{X}'\mathbf{W}_{(m_j-1)}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_{(m_j-1)}Y$, then

$$\mathbf{X}'\mathbf{W}_{(m_j-1)}(Y - \mathbf{X}\beta^{(m_j-1)}) = \mathbf{X}'\mathbf{W}_{(m_j-1)}Y - \mathbf{X}'\mathbf{W}_{(m_j-1)}Y = 0.$$

As a result,

$$
\begin{aligned}
&\|Y-\mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}}-\|Y-\mathbf{X}\beta^{(m_j-1)}\|^2_{\mathbf{W}_{(m_j-1)}}\\
&\quad=\|Y-\mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}}-\|Y-\mathbf{X}\beta^{(m_j-1)}\|^2_{\mathbf{W}_{(m_j-1)}}\\
&\qquad-2(\beta^{(m_j-1)}-\beta^{(m_j)})'\,\mathbf{X}'\mathbf{W}_{(m_j-1)}(Y-\mathbf{X}\beta^{(m_j-1)})\\
&\quad=\|Y-\mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}}+\|Y-\mathbf{X}\beta^{(m_j-1)}\|^2_{\mathbf{W}_{(m_j-1)}}\\
&\qquad-2(Y-\mathbf{X}\beta^{(m_j-1)})'\,\mathbf{W}_{(m_j-1)}(Y-\mathbf{X}\beta^{(m_j-1)})\\
&\qquad-2(\mathbf{X}\beta^{(m_j-1)}-\mathbf{X}\beta^{(m_j)})'\,\mathbf{W}_{(m_j-1)}(Y-\mathbf{X}\beta^{(m_j-1)})\\
&\quad=\|Y-\mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}}+\|Y-\mathbf{X}\beta^{(m_j-1)}\|^2_{\mathbf{W}_{(m_j-1)}}\\
&\qquad-2(Y-\mathbf{X}\beta^{(m_j)})'\,\mathbf{W}_{(m_j-1)}(Y-\mathbf{X}\beta^{(m_j-1)})\\
&\quad=\|(Y-\mathbf{X}\beta^{(m_j)})-(Y-\mathbf{X}\beta^{(m_j-1)})\|^2_{\mathbf{W}_{(m_j-1)}}\\
&\quad=\|\mathbf{X}\beta^{(m_j-1)}-\mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}}.
\end{aligned}
$$

Hence $\|\mathbf{X}\beta^{(m_j-1)}-\mathbf{X}\beta^{(m_j)}\|^2_{\mathbf{W}_{(m_j-1)}}\to 0$. It follows that $\mathbf{X}\beta^{(m_j-1)}-\mathbf{X}\beta^{(m_j)}\to 0$ as $m_j\to\infty$.

Letting $\mathbf{W}=\operatorname{diag}[n_1\mathbf{I}_{n_1}:n_2\mathbf{I}_{n_2}:\cdots:n_k\mathbf{I}_{n_k}]$ and using Lemma 2 of Shi (1994) we have

$$
\begin{aligned}
\|\sigma^{2(m_j)}-\sigma^{2(m_j-1)}\|^2_{\mathbf{W}}&\leqslant\|p^{2(m_j)}-p^{2(m_j-1)}\|^2_{\mathbf{W}}\\
&=\sum_{i=1}^{k} n_i(p_i^{2(m_j)}-p_i^{2(m_j-1)})^2\\
&=\sum_{i=1}^{k}(1/n_i)[\|Y_i-\mathbf{X}_i\beta^{(m_j-1)}\|^2-\|Y_i-\mathbf{X}_i\beta^{(m_j-2)}\|^2]^2.
\end{aligned}
$$

Since $\mathbf{X}\beta^{(m_j-1)}-\mathbf{X}\beta^{(m_j)}\to 0$, then $\|Y_i-\mathbf{X}_i\beta^{(m_j-1)}\|^2-\|Y_i-\mathbf{X}_i\beta^{(m_j-2)}\|^2\to 0$. It follows that $\sigma_i^{2(m_j)}-\sigma_i^{2(m_j-1)}\to 0$ as $m_j\to\infty$ and the lemma is proved.

LEMMA A.3. *Let* $\{\sigma^{2(m)}\}$ *be the sequence of estimated variances from Algorithm* 2.1. *Let* $\{\sigma^{2(m_j)}\}$ *be a subsequence of* $\{\sigma^{2(m)}\}$. *If* $\{\sigma^{2(m_j)}\}$ *converges, then it converges to a favorable point.*

Proof. Let v be the value to which $\{\sigma^{2(m_j)}\}$ componentwise converges. Since $v\in\mathscr{D}$, there exists a subscript set such that v satisfies (5) where $i_{(t+1)}=k$. Let $i_0=0$,

$$
\alpha_{s+1}^{(m_j)}=\min_{i_s+1\leqslant i\leqslant i_{(s+1)}}\{\sigma_i^{2(m_j)}\},\qquad\text{and}\qquad \gamma_{s+1}^{(m_j)}=\max_{i_s+1\leqslant i\leqslant i_{(s+1)}}\{\sigma_i^{2(m_j)}\}
$$

for $s=0, 1, \ldots, t$. Since v satisfies (5), then $\lim_{m_j \to \infty} \alpha_{s+1}^{(m_j)} = \lim_{m_j \to \infty} \gamma_{s+1}^{(m_j)}$. Hence for any $\delta > 0$, there exists an integer M^* such that

$$\max_{s=0, 1, \ldots, t} \{\gamma_{s+1}^{(m^*)} - \alpha_{s+1}^{(m^*)}\} < \delta \tag{17}$$

for every $m^* > M^*$.

From (17) and since $\sigma^{2(m_j)}$ is an isotonic regression of $p^{2(m_j)}$, we may apply a lemma of Barlow *et al.* (1972, p. 34). Therefore for any $\varepsilon > 0$ there exists an integer $M > M^*$ such that

$$\left| \sum_{i=i_s+1}^{i_{(s+1)}} n_i(\sigma_i^{2(m^*)} - p_i^{2(m^*)}) \right| < \varepsilon \tag{18}$$

for every $m^* > M$.

Observe that left hand side of (6) evaluated for v is

$$\sum_{i=i_s+1}^{i_{(s+1)}} n_i v_i - \|Y_i - \mathbf{X}_i(\mathbf{X}'\mathbf{W}_v\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_v Y\|^2$$

$$= \lim_{m_j \to \infty} \sum_{i=i_s+1}^{i_{(s+1)}} n_i\{\sigma_i^{2(m_j)} - (1/n_i)\,\|Y_i - \mathbf{X}_i(\mathbf{X}'\mathbf{W}_{(m_j)}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_{(m_j)}Y\|^2\}$$

$$= \lim_{m_j \to \infty} \sum_{i=i_s+1}^{i_{(s+1)}} n_i\{\sigma_i^{2(m_j)} - (1/n_i)\,\|Y_i - \mathbf{X}_i(\mathbf{X}'\mathbf{W}_{(m_j-1)}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{W}_{(m_j-1)}Y\|^2\}$$

by Lemma A.2

$$= 0 \qquad \text{by (18).}$$

Thus v is a favorable point, and the lemma is proved.

Proof of Theorem 2.2. Let $l_{2m-1} = l(\beta^{(m-1)}, \sigma^{2(m)})$ and $l_{2m} = l(\beta^{(m)}, \sigma^{2(m)})$ for $m \geqslant 1$. Then $\{l_{m^*}\}$ is monotone non-decreasing, by (3) and (4) for $m^* = 1, 2, \ldots$. Let $\hat{\beta}$ and $\hat{\sigma}^2$ be the MLEs for β and σ^2, respectively, in the case where there are no order restrictions on either parameter. Then $l_{m^*} \leqslant l(\hat{\beta}, \hat{\sigma}^2)$ for all $m^* = 1, 2, \ldots$. Hence $\{l_{m^*}\}$ is a bounded monotone non-decreasing sequence. Therefore the sequence converges, and $l_{m^*} - l_{(m^*-1)} \to 0$ as $m^* \to 0$.

Suppose that $\{\sigma^{2(m)}\}$ does not converge. Lemma A1 of Shi and Jiang (1998) shows that for a uniformly bounded sequence $\{y_n\}$ such that $y_n - y_{(n-1)} \to 0$ as $n \to \infty$, if the sequence is not convergent then there are infinitely many accumulation points of the sequence. From Lemma A.1 it is seen that $\{\sigma^{2(m)}\}$ is a bounded sequence. Since $l_{m^*} - l_{(m^*-1)} \to 0$, by the same methods as used in the proof of Lemma A.2, it can be shown that

$\sigma^{2(m)} - \sigma^{2(m-1)} \to 0$. Thus there are infinitely many accumulation points for $\{\sigma^{2(m)}\}$. By Lemma A.3 the accumulation points are all favorable points. Thus if $\{\sigma^{2(m)}\}$ does not converge then there are infinitely many favorable points. However, by assumption there are finitely many favorable points. Therefore, $\{\sigma^{2(m)}\}$ is convergent and by Lemma A.3 $\{\sigma^{2(m)}\}$ converges to a favorable point.

Proof of Theorem 3.1. The complete data can be defined as $\mathcal{Y} = [Y' : \varepsilon']'$. Under the assumption of multivariate normality, then $\mathcal{Y} \sim N(\mu_{\mathcal{Y}}, \Xi)$ where

$$\mu_{\mathcal{Y}} = \begin{pmatrix} \mathbf{X}\beta \\ 0 \end{pmatrix}, \qquad \Xi = \begin{pmatrix} \mathbf{\Psi} & \mathbf{\Sigma} \\ \mathbf{\Sigma} & \mathbf{\Sigma} \end{pmatrix},$$

and $\mathbf{\Psi} = \mathbf{UTU}' + \mathbf{\Sigma}$. Thus the likelihood function is

$$f(\sigma^2; \mathcal{Y}, \beta, \tau^2) = (2\pi)^{-(N+N)/2} \, |\Xi|^{-(1/2)} \exp\{-\tfrac{1}{2}(\mathcal{Y} - \mu_{\mathcal{Y}})' \, \Xi^{-1}(\mathcal{Y} - \mu_{\mathcal{Y}})\}.$$

Recall that

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{vmatrix} = |\mathbf{C}| \cdot |\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}'| \tag{19}$$

(cf. Searle *et al.*, 1992). Since

$$\mathbf{\Psi} - \mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{\Sigma} = \mathbf{\Psi} - \mathbf{\Sigma} = \mathbf{UTU}', \tag{20}$$

it follows that

$$|\Xi| = |\mathbf{\Sigma}| \, |\mathbf{\Psi} - \mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}| = \left(\prod_{i=1}^{k} \sigma_i^{2n_i}\right) |\mathbf{UTU}'|.$$

Additionally, recall that

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{C}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{C}^{-1}\mathbf{B}' \end{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1} \, (\mathbf{I}_N - \mathbf{B}\mathbf{C}^{-1}) \tag{21}$$

(cf. Searle *et al.*, 1992). Again using (20), then

$$\Xi^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{\Sigma}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{I}_N \end{pmatrix} (\mathbf{UTU}')^{-1} \, (\mathbf{I}_N - \mathbf{I}_N).$$

As $(\mathscr{Y}-\mu_{\mathscr{Y}}) = [(Y-\mathbf{X}\beta)' : \varepsilon']'$, it follows that

$$(\mathscr{Y}-\mu_{\mathscr{Y}})'\,\Xi^{-1}(\mathscr{Y}-\mu_{\mathscr{Y}}) = ((Y-\mathbf{X}\beta)'\,\varepsilon')\begin{pmatrix}0 & 0\\ 0 & \mathbf{\Sigma}^{-1}\end{pmatrix}\begin{pmatrix}Y-\mathbf{X}\beta\\ \varepsilon\end{pmatrix}$$

$$+((Y-\mathbf{X}\beta)'\,\varepsilon')\begin{pmatrix}\mathbf{I}_N\\ -\mathbf{I}_N\end{pmatrix}(\mathbf{UTU}')^{-1}\,(\mathbf{I}_N-\mathbf{I}_N)$$

$$\cdot\begin{pmatrix}Y-\mathbf{X}\beta\\ \varepsilon\end{pmatrix}$$

$$=\varepsilon'\mathbf{\Sigma}^{-1}\varepsilon+(Y-\mathbf{X}\beta-\varepsilon)'\,(\mathbf{UTU}')^{-1}\,(Y-\mathbf{X}\beta-\varepsilon).$$

Hence the likelihood function is

$$f(\sigma^2; \mathscr{Y}, \beta, \tau^2)$$

$$=(2\pi)^{-N}\left[\left(\prod_{i=1}^{k}\sigma_i^{2n_i}\right)|\mathbf{UTU}'|\right]^{-(1/2)}$$

$$\cdot\exp\left\{-\tfrac{1}{2}\left[\sum_{i=1}^{k}\sigma_i^{-2}\varepsilon_i'\varepsilon_i+(Y-\mathbf{X}\beta-\varepsilon)'\,(\mathbf{UTU}')^{-1}\,(Y-\mathbf{X}\beta-\varepsilon)\right]\right\},$$

and the log-likelihood is

$$l(\sigma^2; \mathscr{Y}, \beta, \tau^2) = -N\ln(2\pi)-(1/2)\left(\sum_{i=1}^{k}n_i\ln\sigma_i^2\right)-(1/2)\ln|\mathbf{UTU}'|$$

$$-(1/2)\sum_{i=1}^{k}\sigma_i^{-2}\varepsilon_i'\varepsilon_i-(1/2)(Y-\mathbf{X}\beta-\varepsilon)'$$

$$\cdot(\mathbf{UTU}')^{-1}\,(Y-\mathbf{X}\beta-\varepsilon). \tag{22}$$

The maximization step of the EM algorithm finds the maximum likelihood estimators of the unknown parameters as if the complete data were available.

Taking the derivative of (22) with respect to σ_i^2,

$$\frac{\partial l}{\partial\sigma_i^2} = -(1/2)\,n_i(1/\sigma_i^2)+(1/2)[\varepsilon_i'\varepsilon_i/(\sigma_i^2)^2]$$

for $i = 1, \ldots, k$. Setting the derivative equal to 0 and simplifying yields

$$\hat{\sigma}_i^2 = (1/n_i)\,\varepsilon_i'\varepsilon_i \tag{23}$$

for $i = 1, \ldots, k$.

The expectation step of the EM algorithm replaces the unknown quantities in (23) with their conditional expected values, where the conditioning is on the known values.

Recall that if

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}\right),$$

then $E(\mathbf{X}_2 \mid \mathbf{X}_1) = \mu_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \mu_1)$ and $\mathrm{Cov}(\mathbf{X}_2 \mid \mathbf{X}_1) = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$. Observe that

$$\begin{pmatrix} Y \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}\beta \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{\Psi} & \sigma_i^2 \Gamma_i \\ \sigma_i^2 \Gamma'_i & \sigma_i^2 \mathbf{I}_{n_i} \end{pmatrix}\right),$$

where $\Gamma_i = [0 : \cdots : 0 : \mathbf{I}_{n_i} : 0 : \cdots : 0]'$. Thus

$$E(\varepsilon_i \mid Y) = \sigma_i^2 \Gamma'_i \mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta), \qquad \text{and} \qquad \mathrm{Cov}(\varepsilon_i \mid Y) = \sigma_i^2 \mathbf{I}_{n_i} - \sigma_i^4 \Gamma'_i \mathbf{\Psi}^{-1} \Gamma_i$$

for $i = 1, \ldots, k$. Hence

$$\begin{aligned} E(\varepsilon'_i \varepsilon_i \mid Y) &= (E\varepsilon_i \mid Y)' \, (E\varepsilon_i \mid Y) + \mathrm{tr}[\mathrm{Cov}(\varepsilon_i \mid Y)] \\ &= \sigma_i^4 (Y - \mathbf{X}\beta)' \, \mathbf{\Psi}^{-1} \Gamma_i \Gamma'_i \mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta) + n_i \sigma_i^2 - \sigma_i^4 \, \mathrm{tr}(\Gamma'_i \mathbf{\Psi}^{-1} \Gamma_i) \\ &= n_i \sigma_i^2 + \sigma_i^4 \{\mathrm{tr}[\Gamma'_i \mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta)(Y - \mathbf{X}\beta)' \, \mathbf{\Psi}^{-1} \Gamma_i] - \mathrm{tr}(\Gamma'_i \mathbf{\Psi}^{-1} \Gamma_i)\} \\ &= n_i \sigma_i^2 + \sigma_i^4 \, \mathrm{tr}[\mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta)(Y - \mathbf{X}\beta)' \, \mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}]_{ii}, \end{aligned} \tag{24}$$

where $(\mathbf{A})_{ii}$ indicates the (i, i) block of $\mathbf{A}$.

The algorithm begins by selecting some initial estimates for σ_i^2, $i = 1, \ldots, k$. Denote these initial estimates by $\hat{\sigma}_i^{2(0)}$, $i = 1, \ldots, k$. Additionally, let $\hat{\sigma}_i^{2(r)}$, $i = 1, \ldots, k$ indicate the value of the respective parameters at the rth iteration. At the rth iteration of the algorithm, the estimates are updated by substituting the conditional expectations given in Eq. (24) into the maximization equations of (23).

Let $\hat{\mathbf{\Sigma}}^{(r-1)} = \mathrm{diag}[\hat{\sigma}_1^{2(r-1)} \mathbf{I}_{n_1} : \cdots : \hat{\sigma}_k^{2(r-1)} \mathbf{I}_{n_k}]$ and $\hat{\mathbf{\Psi}}^{(r-1)} = \mathbf{UTU}' + \hat{\mathbf{\Sigma}}^{(r-1)}$. Observe that

$$\lambda_{\min}(\mathbf{UTU}' + \hat{\mathbf{\Sigma}}^{(r-1)}) \geqslant \lambda_{\min}(\mathbf{UTU}') + \lambda_{\min}(\hat{\mathbf{\Sigma}}^{(r-1)})$$

(Stewart, 1973, p. 315). Suppose that $\hat{\sigma}_i^{2(r-1)} \geqslant 0$ for all $i = 1, 2, \ldots, k$ and for all r. Then $\lambda_{\min}(\hat{\mathbf{\Sigma}}^{(r-1)}) \geqslant 0$. Since $\lambda_{\min}(\mathbf{UTU}') > 0$, then $\lambda_{\min}(\mathbf{UTU}' + \hat{\mathbf{\Sigma}}^{(r-1)}) > 0$. It follows that $\hat{\mathbf{\Psi}}^{(r-1)}$ is of full rank and invertible. Hence

$$\begin{aligned} \hat{\sigma}_i^{2(r)} = \hat{\sigma}_i^{2(r-1)} &+ (\hat{\sigma}_i^{4(r-1)} / n_i) \\ &\cdot \mathrm{tr}[(\hat{\mathbf{\Psi}}^{(r-1)})^{-1} \, (Y - \mathbf{X}\beta)(Y - \mathbf{X}\beta)' \, (\hat{\mathbf{\Psi}}^{(r-1)})^{-1} - (\hat{\mathbf{\Psi}}^{(r-1)})^{-1}]_{ii}, \end{aligned}$$

for $i = 1, \ldots, k$. Notice that if $\hat{\sigma}_i^{2(0)} \geqslant 0$ then $\hat{\sigma}_i^{2(r)} \geqslant 0$ for all $i = 1, \ldots, k$ and for all r. Thus the theorem is proved.

Proof of Theorem 3.2. The complete data can be defined as $\mathscr{Y} = [Y' : \xi']'$. Under the assumption of multivariate normality, then $\mathscr{Y} \sim N(\mu_{\mathscr{Y}}, \Xi)$ where

$$\mu_{\mathscr{Y}} = \begin{pmatrix} \mathbf{X}\beta \\ 0 \end{pmatrix} \quad \text{and} \quad \Xi = \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{UT} \\ \mathbf{TU}' & \mathbf{T} \end{pmatrix}.$$

Thus the likelihood function is

$$f(\beta, \tau^2; \mathscr{Y}, \sigma^2) = (2\pi)^{-(N+c)/2} \, |\Xi|^{-(1/2)} \exp\{-\tfrac{1}{2}(\mathscr{Y} - \mu_{\mathscr{Y}})' \, \Xi^{-1}(\mathscr{Y} - \mu_{\mathscr{Y}})\},$$

where $c = \sum_{i=1}^{q} c_i$.

Since

$$\boldsymbol{\Psi} - \mathbf{UTT}^{-1}\mathbf{TU}' = \boldsymbol{\Psi} - \mathbf{UTU}' = \boldsymbol{\Sigma}, \tag{25}$$

it follows from (19) that

$$|\Xi| = |T| \, |\boldsymbol{\Psi} - \mathbf{UTT}^{-1}\mathbf{TU}'| = \left(\prod_{i=1}^{q} \tau_i^{2c_i}\right) |\boldsymbol{\Sigma}|.$$

Using (21) and (25), then

$$\Xi^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & T^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{U}' \end{pmatrix} \boldsymbol{\Sigma}^{-1}(\mathbf{I}_N - \mathbf{U}).$$

As $(\mathscr{Y} - \mu_{\mathscr{Y}}) = [(Y - \mathbf{X}\beta)' : \xi']'$, it follows that

$$\begin{aligned}(\mathscr{Y} - \mu_{\mathscr{Y}})' \, \Xi^{-1}(\mathscr{Y} - \mu_{\mathscr{Y}}) &= ((Y - \mathbf{X}\beta)' \, \xi') \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{T}^{-1} \end{pmatrix} \begin{pmatrix} Y - \mathbf{X}\beta \\ \xi \end{pmatrix} \\ &\quad + ((Y - \mathbf{X}\beta)' \, \xi') \begin{pmatrix} \mathbf{I}_N \\ -\mathbf{U}' \end{pmatrix} \boldsymbol{\Sigma}^{-1}(\mathbf{I}_N - \mathbf{U}) \cdot \begin{pmatrix} Y - \mathbf{X}\beta \\ \xi \end{pmatrix} \\ &= \xi' \mathbf{T}^{-1} \xi + (Y - \mathbf{X}\beta - \mathbf{U}\xi)' \, \boldsymbol{\Sigma}^{-1}(Y - \mathbf{X}\beta - \mathbf{U}\xi).\end{aligned}$$

Hence the likelihood function is

$$\begin{aligned} &f(\beta, \tau^2; \mathscr{Y}, \sigma^2) \\ &\quad = (2\pi)^{-(N+c)/2} \left[\left(\prod_{i=1}^{q} \tau_i^{2c_i}\right) |\boldsymbol{\Sigma}|\right]^{-(1/2)} \\ &\qquad \cdot \exp\left\{-\tfrac{1}{2}\left[\sum_{i=1}^{q} \tau_i^{-2} \xi_i' \xi_i + (Y - \mathbf{X}\beta - \mathbf{U}\xi)' \, \boldsymbol{\Sigma}^{-1}(Y - \mathbf{X}\beta - \mathbf{U}\xi)\right]\right\},\end{aligned}$$

and the log-likelihood is

$$l(\beta, \tau^2; \mathscr{Y}, \sigma^2) = [-(N+c)/2] \ln(2\pi) - (1/2)\left(\sum_{i=1}^{q} c_i \ln \tau_i^2\right)$$
$$-(1/2) \ln \boldsymbol{\Sigma} - (1/2) \sum_{i=1}^{q} \tau_i^{-2} \xi_i' \xi_i$$
$$-(1/2)(Y - \mathbf{X}\beta - \mathbf{U}\xi)' \boldsymbol{\Sigma}^{-1} (Y - \mathbf{X}\beta - \mathbf{U}\xi). \tag{26}$$

The maximization step of the EM algorithm finds the maximum likelihood estimators of the unknown parameters as if the complete data were available.

Taking the derivative of (26) with respect to β,

$$\frac{\partial l}{\partial \beta} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}(Y - \mathbf{U}\xi) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\beta$$

as $\partial(b'a)/\partial b = a$ and $\partial(b'\mathbf{C}b)/\partial b = 2\mathbf{C}b$ for vectors a, b and matrix $\mathbf{C}$. Setting the derivative equal to 0 and simplifying yields

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\beta} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}(Y - \mathbf{U}\xi). \tag{27}$$

Since $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ is invertible, then $\hat{\beta} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}(Y - \mathbf{U}\xi)$. Taking the derivative of (26) with respect to τ_i^2,

$$\frac{\partial l}{\partial \tau_i^2} = -(1/2)\, c_i (1/\tau_i^2) + (1/2)[\xi_i'\xi_i/(\tau_i^2)^2]$$

for $i = 1, \ldots, q$. Setting the derivative equal to 0 and simplifying yields

$$\hat{\tau}_i^2 = (1/c_i)\, \xi_i'\xi_i \tag{28}$$

for $i = 1, \ldots, q$.

The expectation step of the EM algorithm replaces the unknown quantities in (27) and (28) with their conditional expected values, where the conditioning is on the known values. In Eqs. (27) and (28), the unknown quantities are the ξ_i. Observe that

$$\begin{pmatrix} Y \\ \xi_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}\beta \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Psi} & \tau_i^2 \mathbf{U}_i \\ \tau_i^2 \mathbf{U}_i' & \tau_i^2 \mathbf{I}_{c_i} \end{pmatrix}\right),$$

where $\boldsymbol{\Psi} = \mathbf{U}\mathbf{T}\mathbf{U}' + \boldsymbol{\Sigma}$. Thus

$$E(\xi_i \mid Y) = \tau_i^2 \mathbf{U}_i' \boldsymbol{\Psi}^{-1}(Y - \mathbf{X}\beta), \qquad \text{and} \qquad \operatorname{Cov}(\xi_i \mid Y) = \tau_i^2 \mathbf{I}_{c_i} - \tau_i^4 \mathbf{U}_i' \boldsymbol{\Psi}^{-1} \mathbf{U}_i$$

for $i = 1, \ldots, q$. Hence

$$E(\xi \mid Y)' = \mathbf{T}\mathbf{U}'\mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta), \qquad \text{and} \tag{29}$$

$$\begin{aligned} E(\xi_i'\xi_i \mid Y) &= (E\xi_i \mid Y)'\,(E\xi_i \mid Y) + \operatorname{tr}[\operatorname{Cov}(\xi_i \mid Y)] \\ &= \tau_i^4 (Y - \mathbf{X}\beta)'\,\mathbf{\Psi}^{-1}\mathbf{U}_i\mathbf{U}_i'\mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta) + c_i\tau_i^2 - \tau_i^4 \operatorname{tr}(\mathbf{U}_i'\mathbf{\Psi}^{-1}\mathbf{U}_i) \\ &= c_i\tau_i^2 + \tau_i^4 \{\operatorname{tr}[\mathbf{U}_i'\mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta)(Y - \mathbf{X}\beta)'\,\mathbf{\Psi}^{-1}\mathbf{U}_i] - \operatorname{tr}(\mathbf{U}_i'\mathbf{\Psi}^{-1}\mathbf{U}_i)\} \\ &= c_i\tau_i^2 + \tau_i^4 \operatorname{tr}\{\mathbf{U}_i'[\mathbf{\Psi}^{-1}(Y - \mathbf{X}\beta)(Y - \mathbf{X}\beta)'\,\mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}]\,\mathbf{U}_i\}. \end{aligned} \tag{30}$$

The algorithm begins by selecting some initial estimates for β and τ_i^2, $i = 1, \ldots, q$. Denote these initial estimates by $\hat{\beta}^{(0)}$ and $\hat{\tau}_i^{2(0)}$, $i = 1, \ldots, q$ respectively. Additionally, let $\hat{\beta}^{(r)}$ and $\hat{\tau}_i^{2(r)}$, $i = 1, \ldots, q$ indicate the value of the respective parameters at the rth iteration. At the rth iteration of the algorithm, the estimates are updated by substituting the conditional expectations given in Eqs. (29) and (30) into the maximization equations of (27) and (28), respectively.

Let $\hat{\mathbf{T}}^{(r-1)} = \operatorname{diag}[\hat{\tau}_1^{2(r-1)}\mathbf{I}_{c_1} : \cdots : \hat{\tau}_q^{2(r-1)}\mathbf{I}_{c_q}]$ and $\hat{\mathbf{\Psi}}^{(r-1)} = \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \mathbf{\Sigma}$. Observe that

$$\lambda_{\min}(\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \mathbf{\Sigma}) \geqslant \lambda_{\min}(\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}') + \lambda_{\min}(\mathbf{\Sigma})$$

(Stewart, 1973, p. 315). Suppose that $\hat{\tau}_i^{2(r-1)} \geqslant 0$ for all $i = 1, 2, \ldots, q$ and for all r. Then $\lambda_{\min}(\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}') \geqslant 0$. Since $\lambda_{\min}(\mathbf{\Sigma}) > 0$, then $\lambda_{\min}(\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \mathbf{\Sigma}) > 0$ and $\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \mathbf{\Sigma}$ is of full rank. Thus when $\hat{\tau}_i^{2(r-1)} \geqslant 0$ for all i and r, combining (27) and (29) gives

$$\begin{aligned} \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\hat{\beta}^{(r)} &= \mathbf{X}'\mathbf{\Sigma}^{-1}[Y - \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}'(\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \mathbf{\Sigma})^{-1}\,(Y - \mathbf{X}\hat{\beta}^{(r-1)})] \\ &= \mathbf{X}'\mathbf{\Sigma}^{-1}Y + \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{\Sigma}(\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \mathbf{\Sigma})^{-1}\,(Y - \mathbf{X}\hat{\beta}^{(r-1)}) \\ &\quad - \mathbf{X}'\mathbf{\Sigma}^{-1}(\mathbf{\Sigma} + \mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}')(\mathbf{U}\hat{\mathbf{T}}^{(r-1)}\mathbf{U}' + \mathbf{\Sigma})^{-1}\,(Y - \mathbf{X}\hat{\beta}^{(r-1)}) \\ &= \mathbf{X}'\mathbf{\Sigma}^{-1}Y + \mathbf{X}'(\hat{\mathbf{\Psi}}^{(r-1)})^{-1}\,(Y - \mathbf{X}\hat{\beta}^{(r-1)}) - \mathbf{X}'\mathbf{\Sigma}^{-1}(Y - \mathbf{X}\hat{\beta}^{(r-1)}) \\ &= \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\hat{\beta}^{(r-1)} + \mathbf{X}'(\hat{\mathbf{\Psi}}^{(r-1)})^{-1}\,(Y - \mathbf{X}\hat{\beta}^{(r-1)}). \end{aligned}$$

Since $\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}$ is of full rank, then

$$\hat{\beta}^{(r)} = \hat{\beta}^{(r-1)} + (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'(\hat{\mathbf{\Psi}}^{(r-1)})^{-1}\,(Y - \mathbf{X}\hat{\beta}^{(r-1)}).$$

Additionally, combining (28) and (30) yields

$$\begin{aligned} \hat{\tau}_i^{2(r)} = \hat{\tau}_i^{2(r-1)} + (\hat{\tau}_i^{4(r-1)}/c_i) \operatorname{tr} \mathbf{U}_i'[(\hat{\mathbf{\Psi}}^{(r-1)})^{-1}\,(Y - \mathbf{X}\hat{\beta}^{(r-1)}) \\ \cdot (Y - \mathbf{X}\hat{\beta}^{(r-1)})'\,(\hat{\mathbf{\Psi}}^{(r-1)})^{-1} - (\hat{\mathbf{\Psi}}^{(r-1)})^{-1}]\,\mathbf{U}_i, \end{aligned}$$

for $i = 1, \ldots, q$. Note that if $\hat{\tau}_i^{2(0)} \geqslant 0$ then $\hat{\tau}_i^{2(r-1)} \geqslant 0$ for all $i = 1, 2, \ldots, q$ and for all r. Thus the theorem is proved.

ACKNOWLEDGMENTS

We thank Professor Don Richards, Department of Statistics, University of Virginia, for introducing us to the field of algebraic geometry and for providing us with the reference of Sturmfels (1998). We thank Professor Sturmfels, Department of Mathematics, University of California at Berkeley, and Professor Amnon Neeman, Department of Mathematics, Australian National University, for pointing us to a collection of papers in algebraic geometry which relate to our problem. We thank Professor Sturmfels for showing us how his results can be used in our context. The second author is very grateful to late Professor P. Prabhakara Rao for all the discussions on the analysis of groups of independent experiments and their applications to fertilizer trials which motivated this work.

REFERENCES

R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, "Statistical Inference under Order Restrictions," 3rd ed., Wiley, New York, 1972.

A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. Royal Statist. Soc. Ser. B.* **39** (1977), 1–38.

J. T. G. Hwang and S. D. Peddada, Confidence interval estimation subject to order restrictions, *Ann. Statist.* **22** (1994), 67–93.

B. Huber and B. Sturmfels, A polyhedral method for solving sparse polynomial systems, *Math. Comp.* **64** (1995), 1541–1555.

R. K. Khosla, P. P. Rao, and M. N. Das, A note on the study of the experimental errors in groups of agricultural field experiments conducted in different years, *J. Indian Soc. Agricultural Statist.* **31** (1979), 65–68.

P. Pedersen and B. Sturmfels, Product formulas for resultants and Chow forms, *Math. Z.* **214** (1993), 377–396.

T. Robertson, F. T. Wright, and R. L. Dykstra, "Order Restricted Statistical Inference," Wiley, New York, 1988.

C. R. Rao, V. K. Srivastava, and H. Toutenburg, Pitman nearness comparisons of Stein-type estimators for regression coefficients in replicated experiments, *Statist. Papers* **39** (1998), 61–74.

P. P. Rao, P. N. Bhargava, and J. K. Kapoor, "Analysis of Repeated Experiments," Technical Report, Indian Agricultural Statistics Research Institute, New Delhi, India, 1987.

P. S. R. S. Rao and Y. P. Chaubey, Three modifications of the principle of the MINQUE, *Comm. Statist. Theory. Methods. A* **7** (1978), 767–778.

S. R. Searle, G. Casella and C. E. McCulloch, "Variance Components," Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, Wiley, New York, 1992.

N. Shi, Maximum likelihood estimation of means and variances from normal populations under simultaneous order restrictions, *J. Multivariate. Anal.* **50** (1994), 282–293.

N. Shi and H. Jiang, Maximum likelihood estimation of isotonic normal means with unknown variances, *J. Multivariate. Anal.* **64** (1998), 183–195.

V. K. Srivastava and H. Toutenburg, Application of Stein-type estimation in combining regression estimates from replicated experiments, *Statist. Papers* **35** (1994), 101–112.

G. W. Stewart, "Introduction to Matrix Computations," Computer Science and Applied Mathematics, Academic Press, New York, 1973.

B. Sturmfels, On the Newton polytope of the resultant, *J. Algebraic Combin.* **3** (1994), 207–236.

B. Sturmfels, Polynomial equations and convex polytopes, *Amer. Math. Monthly* **105** (1998), 907–922.